ROCZNIKI POLSKIEGO TOWARZYSTWA MATEMATYCZNEGO
Seria III: MATEMATYKA STOSOWANA XVI (1980)

Ministerstwo Nauki i Szkolnictwa Wyższego

GRAŻYNA HILLE (Warszawa)

Metoda SUBQ w programowaniu kwadratowym

(*Praca przyjęta do druku 21.3.1978*)

Streszczenie. W programowaniu liniowym znana jest powszechnie metoda SUB (jest to skrót nazwy angielskiej: The Simple Upper Bound Method). Jest ona modyfikacją sympleksowej metody znajdującą minimum liniowej funkcji celu w pewnym zbiorze dopuszczalnym, przy czym na zmienne x nałożone są dodatkowe warunki postaci $\beta \leqslant x \leqslant \alpha$.

W niniejszej pracy zaprezentowana jest metoda, która rozwiązuje zadanie minimalizacji kwadratowej, wypukłej funkcji celu w zbiorze dopuszczalnym określonym przez warunki postaci $\beta \leqslant x \leqslant \alpha$. Jest ona modyfikacją algorytmu Lemkego [3] dla zadań programowania kwadratowego. Wobec specyficznej postaci warunków definiujących zbiór dopuszczalny metoda ta jest, w pewnym sensie, odpowiednikiem metody SUB dla programowania liniowego. Dlatego przedstawiony algorytm będziemy nazywali algorytmem SUBQ (The Simple Upper Bound Algorithm for Quadratic Programming).

W § 2 pracy podany jest opis metody oparty głównie na jej interpretacji geometrycznej. W § 3 przedstawiony jest algorytm. § 4 zawiera dowód zbieżności metody.

1. Wstęp. W praktyce często natrafiamy na zadania programowania kwadratowego o następującej postaci:

$$\min\{f(x) = d^T x + \tfrac{1}{2} x^T D x\colon \beta \leqslant x \leqslant \alpha\}, \tag{1.1}$$

gdzie $x \in E^n$, d, β, $\alpha \in E^n$, D jest macierzą symetryczną, nieujemnie określoną.

Nie zmniejszając ogólności zakładamy, że $\beta < \alpha$(1).

Warunki $\beta \leqslant x \leqslant \alpha$, określające zbiór dopuszczalny, można oczywiście sprowadzić do postaci standardowej i zadanie (1.1) rozwiązywać jedną ze znanych metod programowania kwadratowego, jak np. Beale'a, Lemkego, Wolfe'a itp. Postępowanie takie nie uwzględnia jednak specyficznego typu ograniczeń.

W niniejszej pracy przedstawiony jest algorytm, który wykorzystuje szczególną postać warunków określających zbiór dopuszczalny dla zadania (1.1). Nazywać go będziemy, z przyczyn podanych w streszczeniu, algorytmem SUBQ.

Punkt x^k jest punktem optymalnym dla zadania (1.1) wtedy i tylko wtedy, gdy istnieją współczynniki Lagrange'a

$$\lambda^k = \begin{bmatrix} \lambda_\beta^k \\ \lambda_\alpha^k \end{bmatrix}, \qquad \lambda_\beta^k \geqslant 0 \in E^n,\ \lambda_\alpha^k \leqslant 0 \in E^n,$$

(1) Dla wektorów $a, b \in E^n$ określone są relacje: $a \leqslant b \equiv a_i \leqslant b_i$, $i = 1, \ldots, n$, $a < b \equiv a_i < b_i$, $i = 1, \ldots, n$.

spełniające warunki Kuhna–Tuckera:

$$(1.2)\qquad \begin{cases} z^k-\lambda_\beta^k-\lambda_\alpha^k = 0; \ \beta \leqslant x^k \leqslant \alpha; \\ \lambda_\beta^{kT}(x^k-\beta) = 0; \ \lambda_\alpha^{kT}(x^k-\alpha) = 0; \end{cases}$$

gdzie $z^k \overset{\mathrm{df}}{=} \nabla f(x)|_{x=x^k} = d+Dx^k$.

Idea metody SUBQ polega na wyznaczeniu skończonego ciągu $\{(x^k, z^k, \lambda^k)\}$ takiego, aby przy spełnionym zawsze warunku $\beta \leqslant x^k \leqslant \alpha$, pozostałe warunki Kuhna–Tuckera były „coraz mniej naruszane".

Algorytm SUBQ adaptuje w tym celu pomysł Lemkego [3], na którym oparta jest metoda rozwiązująca zadanie programowania kwadratowego postaci

$$(1.3)\qquad \min\{f(x) = d^Tx+\tfrac{1}{2}x^TDx\colon x \geqslant 0\},$$

gdzie $x, d \in E^n$, D jest macierzą symetryczną, nieujemnie określoną.

Założenie o istnieniu skończonego optimum funkcji celu, tak istotne dla zbieżności algorytmu Lemkego, jest tutaj automatycznie spełnione, ponieważ zbiór dopuszczalny $\mathscr{X} \overset{\mathrm{df}}{=} \{x \in E^n\colon \beta \leqslant x \leqslant \alpha\}$ jest zwarty a funkcja celu jest ciągła.

W opracowaniu tym będziemy odwoływać się do pracy [2], w której przedstawiony jest opis oraz interpretacja analityczna i geometryczna metody Lemkego.

2. Opis metody. Rozważmy zbiór $\{x \in E^n\colon a_i^Tx \leqslant b_i, i = 1, \ldots, m\}$. Jeśli x^* jest jego elementem, to warunek $a_j^Tx \leqslant b_j$ nazywamy *istotnym* w punkcie x^* wtedy i tylko wtedy, gdy $a_j^Tx^* = b_j$. Układ warunków $a_{j_1}^Tx \leqslant b_{j_1}, \ldots, a_{j_k}^Tx \leqslant b_{j_k}$ nazywamy układem *liniowo niezależnym* wtedy i tylko wtedy gdy $\{a_{j_1}, \ldots, a_{j_k}\}$ jest liniowo niezależnym zbiorem wektorów. Punkt dopuszczalny $x^* \in \{x \in E^n\colon a_i^Tx \leqslant b_i, i = 1, \ldots, m\}$ jest *ekstremalny* wtedy i tylko wtedy, gdy odpowiada mu układ przynajmniej n warunków istotnych, wśród których jest dokładnie n warunków liniowo niezależnych. Wektory a_j odpowiadające tym warunkom tworzą więc bazę przestrzeni E^n. Zatem, z każdym punktem ekstremalnym x związana jest macierz bazowa (będziemy ją też nazywać „bazą") $P = (p_1, \ldots, p_n)$, której kolumny $p_i = a_{j_i}$ $(i = 1, \ldots, n)$ odpowiadają układowi n liniowo niezależnych warunków istotnych w punkcie x.

W programowaniu kwadratowym punkt optymalny nie musi być ekstremalny. Aby jednak i tu operować pojęciem bazy związanej z punktem dopuszczalnym x, wprowadza się odpowiednią liczbę pomocniczych warunków, które w punkcie x są istotne. Wraz z pozostałymi warunkami istotnymi w x tworzą one układ przynajmniej n warunków istotnych w tym punkcie, przy czym dokładnie n z nich jest liniowo niezależnych.

Opiszemy teraz k-tą iterację metody SUBQ.

Przyjmijmy za punkt startowy algorytmu punkt $x^0 = \beta$. Bazą startową jest $P^0 = I$. Niech punktowi x^k odpowiada baza P^k o kolumnach p_i^k $(i = 1, \ldots, n)$ oraz baza odwrotna (odwrotność bazy P^k):

$$(P^k)^{-1} = \begin{bmatrix} \pi_1^{kT} \\ \vdots \\ \pi_n^{kT} \end{bmatrix},$$

gdzie π_i^{kT} jest i-tym wierszem $(P^k)^{-1}$.

Dla punktu x^k określmy podzbiory $R_\beta^k = R_\beta(x^k)$ oraz $R_\alpha^k = R_\alpha(x^k)$ zbioru $\{1, \ldots, n\}$ takie, że

$$R_\beta^k \stackrel{\mathrm{df}}{=} \{i \in \{1, \ldots, n\}: (\exists j_i \in \{1, \ldots, n\}) p_i^k = e_{j_i} \wedge e_{j_i}^T x^k = \beta_{j_i}\},$$

$$R_\alpha^k \stackrel{\mathrm{df}}{=} \{i \in \{1, \ldots, n\}: (\exists j_i \in \{1, \ldots, n\}) p_i^k = e_{j_i} \wedge e_{j_i}^T x^k = \alpha_{j_i}\}.$$

Wprost z określenia tych zbiorów widać, że

$$R_\beta^k \cap R_\alpha^k = \emptyset.$$

Niech $R^k = R(x^k) \stackrel{\mathrm{df}}{=} R_\beta^k \cup R_\alpha^k$. Będziemy nazywać kolumnę p_i^k kolumną *ograniczoną dolną* (*górną*) w punkcie x^k wtedy i tylko wtedy, gdy $i \in R_\beta^k$ ($i \in R_\alpha^k$).

Kolumna p_i^k będzie nazywana kolumną *ograniczoną* w punkcie x^k wtedy i tylko wtedy, gdy będzie kolumną ograniczoną dolną lub górną w punkcie x^k. Pozostałe kolumny bazy P^k nazywane będą kolumnami *wolnymi* w punkcie x^k.

Podobnie jak w algorytmie Lemkego [2], każdej iteracji odpowiada zmiana bazy P^k. Usuwamy z niej kolumnę p_r^k a w zamian wprowadzamy kolumnę p_r^{k+1}. Indeks r wyznaczamy licząc składowe w_i^k wektora w^k, które są pochodnymi kierunkowymi funkcji celu w punkcie x^k w kierunkach π_i^k:

$$w_i^k \stackrel{\mathrm{df}}{=} f'(x^k, \pi_i^k) = [\nabla f(x)]^T|_{x=x^k} \cdot \pi_i^k = z^{kT}\pi_i^k.$$

Stąd $w^k = (P^k)^{-1} z^k$.

Jeśli istnieje r takie, że $w_r^k \neq 0$, to funkcja celu maleje, gdy z punktu dopuszczalnego x^k przesuwamy się do nowego punktu dopuszczalnego x^{k+1} wzdłuż prostej $x^k - \theta\pi_r^k$, przy czym $\theta \leqslant 0$, gdy $w_r^k < 0$, lub $\theta \geqslant 0$, gdy $w_r^k > 0$. Każdemu punktowi prostej $x^k - \theta\pi_r^k$ odpowiada pewna wartość $w_r^k = w_r^k(\theta) = \pi_r^{kT} z^k(\theta)$, gdzie $z^k(\theta) = d + D(x^k - \theta\pi_r^k)$.

Nowe $x^{k+1} = x^k - \theta^k\pi_r^k$ jest punktem, w którym w_r^k jako funkcja θ staje się zerem (tj. osiągane zostaje minimum bezwarunkowe funkcji celu na prostej $x^k - \theta\pi_r^k$) lub w którym osiągany jest brzeg obszaru dopuszczalnego (tj. osiągane zostaje minimum warunkowe funkcji celu na prostej $x^k - \theta\pi_r^k$).

W algorytmie SUBQ warunkami pomocniczymi są zatem warunki postaci $w_r^k \leqslant 0$ lub $w_r^k \geqslant 0$, które uniemożliwiają dotarcie wzdłuż prostej $x^k - \theta\pi_r^k$ do punktu, w którym funkcja celu zaczyna rosnąć.

Jeśli oznaczymy $q^k \stackrel{\mathrm{df}}{=} D\pi_r^k$, to $w_r^k = w_r^k(\theta) = \pi_r^{kT} d + q^{kT}(x^k - \theta\pi_r^k)$. Stąd warunek pomocniczy $w_r^k \leqslant 0$ ($w_r^k \geqslant 0$) jest istotny w punkcie x^{k+1} wtedy i tylko wtedy, gdy

$$q^{kT} x^{k+1} = -\pi_r^{kT} d. \tag{2.1}$$

Ponadto w nowym punkcie x^{k+1} zachodzi:

$$p_r^{kT} x^{k+1} = p_r^{kT}(x^k - \theta^k\pi_r^k) = p_r^{kT} x^k - \theta^k p_r^{kT}\pi_r^k = p_r^{kT} x^k - \theta^k. \tag{2.2}$$

Jeśli $\theta^k \neq 0$, to warunek odpowiadający kolumnie p_r^k nie jest istotny w punkcie x^{k+1}.

W punkcie x^{k+1} zmieniamy bazę P^k. Usuwamy z niej kolumnę p_r^k (wobec (2.2)) a w zamian wprowadzamy p_r^{k+1}. Kolumna p_r^{k+1} jest kolumną wolną q^k, gdy w punk-

cie x^{k+1} istotny jest warunek pomocniczy (tj. zachodzi zależność (2.1)), lub kolumną ograniczoną (dolną lub górną) e_l, gdy w punkcie x^{k+1} istotny jest któryś z warunków wyjściowych taki, że $e_l^T x^{k+1} = \beta_l$ lub $e_l^T x^{k+1} = \alpha_l$ oraz $e_l \notin P^k$.

Wprowadzenie do bazy kolumny wolnej q^k odpowiada przypadkowi, gdy minimum bezwarunkowe funkcji celu na prostej $x^k - \theta\pi_r^k$ jest przyjmowane wewnątrz zbioru dopuszczalnego $\mathscr{X}$. Natomiast wprowadzenie do bazy kolumny ograniczonej e_l odpowiada przypadkowi, gdy minimum bezwarunkowe funkcji celu na prostej $x^k - \theta\pi_r^k$ jest przyjmowane na brzegu lub poza zbiorem $\mathscr{X}$.

Wyznaczmy teraz wartość θ_0^k parametru θ, dla której funkcja celu przyjmuje minimum bezwarunkowe na prostej $x^k - \theta\pi_r^k$, oraz wartość θ^k parametru θ, dla której funkcja celu osiąga minimum warunkowe na prostej $x^k - \theta\pi_r^k$ w zbiorze dopuszczalnym $\mathscr{X}$.

Podobnie jak w punkcie 3 pracy [2], wprowadźmy

$$\varphi(\theta) \overset{\mathrm{df}}{=} f(x^k - \theta\pi_r^k) = f(x^k) - \theta w_r^k + \tfrac{1}{2}\theta^2 q^{kT}\pi_r^k.$$

Mamy

$$\varphi(0) = f(x^k) \quad \text{oraz} \quad \varphi'(\theta) = -w_r^k + \theta q^{kT}\pi_r^k.$$

θ_0^k jest zatem rozwiązaniem równania $\varphi'(\theta) = 0$, czyli $\theta_0^k = \dfrac{w_r^k}{q^{kT}\pi_r^k} \neq 0$, gdy $q^k \neq 0$ (z założenia $w_r^k \neq 0$). Przypadek $q^k = 0$ rozpatrzymy później.

Niech t_0^k oznacza największy (w sensie supremum) mnożnik rzeczywisty, dla którego $\beta \leqslant x^k - t_0^k\theta_0^k\pi_r^k \leqslant \alpha$. Łatwo zauważyć, że

$$t_0^k = \begin{cases} \min\{t_0^{k\beta}, t_0^{k\alpha}\}, & \text{gdy} \quad \exists j\ \theta_0^k\pi_{rj}^k > 0 \wedge \exists j\ \theta_0^k\pi_{rj}^k < 0, \\ t_0^{k\beta}, & \text{gdy} \quad \exists j\ \theta_0^k\pi_{rj}^k > 0 \wedge \exists j\ \theta_0^k\pi_{rj}^k \geqslant 0, \\ t_0^{k\alpha}, & \text{gdy} \quad \exists j\ \theta_0^k\pi_{rj}^k < 0 \wedge \exists j\ \theta_0^k\pi_{rj}^k \leqslant 0, \\ +\infty, & \text{gdy} \quad \theta_0^k\pi_r^k = 0, \end{cases} \tag{2.3}$$

gdzie

$$t_0^{k\beta} = \frac{x_l^k - \beta_l^k}{\theta_0^k\pi_{rl}^k} = \min_j \left\{\frac{x_j^k - \beta_j^k}{\theta_0^k\pi_{rj}^k} : \theta_0^k\pi_{rj}^k > 0\right\}, \tag{2.4}$$

$$t_0^{k\alpha} = \frac{x_l^k - \alpha_l^k}{\theta_0^k\pi_{rl}^k} = \min_j \left\{\frac{x_j^k - \alpha_j^k}{\theta_0^k\pi_{rj}^k} : \theta_0^k\pi_{rj}^k < 0\right\} \tag{2.5}$$

i że wówczas

$$\theta^k = \begin{cases} \theta_0^k, & \text{jeśli} \quad t_0^k > 1, \\ t_0^k\theta_0^k, & \text{jeśli} \quad 0 \leqslant t_0^k \leqslant 1. \end{cases}$$

Z powyższych formuł wynikają następujące własności:

(i) funkcja φ przyjmuje minimum warunkowe na zbiorze

$$\{\theta\colon \beta \leqslant x^k - \theta\pi_r^k \leqslant \alpha\} \quad \text{w punkcie } \theta = \theta^k;$$

(ii) jeśli $t_0^k > 1$, to $\theta^k = \theta_0^k$, czyli minimum bezwarunkowe φ leży wewnątrz zbioru dopuszczalnego;

(iii) jeśli $t_0^k = 1$, to $\theta^k = \theta_0^k$ i minimum bezwarunkowe φ leży na brzegu zbioru dopuszczalnego;

(iv) jeśli $0 < t_0^k < 1$, to $\theta^k = t_0^k\theta_0^k$ i minimum bezwarunkowe φ leży poza zbiorem dopuszczalnym a warunkowe na jego brzegu: jeśli $t_0^k = t_0^{k\beta}$, to punkt x^{k+1} leży na l-tym ograniczeniu dolnym, jeśli $t_0^k = t_0^{k\alpha}$, to punkt x^{k+1} leży na l-tym ograniczeniu górnym;

(v) jeśli $t_0^k = 0$, to $\theta^k = 0$ i kierunek π_r^k wyprowadza poza zbiór dopuszczalny (kierunek nie jest dopuszczalny); $x^{k+1} = x^k$.

Rozpatrzmy teraz przypadek, gdy $q^k = 0$. Wówczas $\varphi(\theta) = f(x^k) - \theta w_r^k$. Każde niezerowe θ takie, że $\operatorname{sign}\theta = \operatorname{sign} w_r^k$, zmniejsza wartość funkcji celu. Zatem, jeśli posuwamy się z punktu x^k wzdłuż prostej $x^k - \theta\pi_r^k$, to dla pewnego θ takiego, że $\operatorname{sign}\theta = \operatorname{sign} w_r^k$, osiągniemy brzeg obszaru dopuszczalnego. Jeśli t_0^k wyznaczone jest jak dla przypadku gdy $q^k \neq 0$, to $\theta^k = t_0^k\theta_0^k$, gdzie $\theta_0^k = \operatorname{sign} w_r^k$. ($\theta_0^k$ nie wyznacza teraz punktu, w którym funkcja celu przyjmuje minimum bezwarunkowe na prostej $x^k - \theta\pi_r^k$.)

Definicja. Niech zapis $i \notin R^k$ oznacza, że $i \in \{1, \ldots, n\} - R^k$. Określmy zbiory indeksów $\mathscr{K}_{\mathrm{I}}^k \overset{\mathrm{df}}{=} \{i \notin R^k\colon w_i^k \neq 0\}$, $\mathscr{K}_{\mathrm{II}}^k \overset{\mathrm{df}}{=} \{i \in R_\beta^k\colon w_i^k < 0\}$, $\mathscr{K}_{\mathrm{III}}^k \overset{\mathrm{df}}{=} \{i \in R_\alpha^k\colon w_i^k > 0\}$ oraz $\mathscr{K}^k \overset{\mathrm{df}}{=} \mathscr{K}_{\mathrm{I}}^k \cup \mathscr{K}_{\mathrm{II}}^k \cup \mathscr{K}_{\mathrm{III}}^k$. Wówczas $\mathscr{K}_{\mathrm{II}}^k \cap \mathscr{K}_{\mathrm{III}}^k = \emptyset$, bo z założenia $\beta < \alpha$.

Z analitycznego punktu widzenia wybór indeksu r jest determinowany przez 3 przypadki:

(I) $\mathscr{K}_{\mathrm{I}}^k \neq \emptyset$;

(II) $\mathscr{K}_{\mathrm{I}}^k = \emptyset \wedge \mathscr{K}_{\mathrm{II}}^k \neq \emptyset$;

(III) $\mathscr{K}_{\mathrm{I}}^k = \emptyset \wedge \mathscr{K}_{\mathrm{III}}^k \neq \emptyset$.

Stanowią one podstawę do określenia poniższych strategii wyboru indeksu r:

A (a) jeśli $\mathscr{K}_{\mathrm{I}}^k \neq \emptyset$, to r jest indeksem składowej w_r^k takiej, że $|w_r^k| = \max\limits_{i \in \mathscr{K}_{\mathrm{I}}^k} |w_i^k|$;

(b) jeśli $\mathscr{K}_{\mathrm{I}}^k = \emptyset$ oraz $\mathscr{K}_{\mathrm{II}}^k \neq \emptyset$, to r odpowiada składowej w_r^k takiej, że $|w_r^k| = \max\limits_{i \in \mathscr{K}_{\mathrm{II}}^k} |w_i^k|$;

(c) jeśli $\mathscr{K}_{\mathrm{I}}^k = \emptyset$, $\mathscr{K}_{\mathrm{II}}^k = \emptyset$ oraz $\mathscr{K}_{\mathrm{III}}^k \neq \emptyset$, to r odpowiada składowej w_r^k takiej, że $|w_r^k| = \max\limits_{i \in \mathscr{K}_{\mathrm{III}}^k} |w_i^k|$.

B (a) Tak samo jak A (a);

(b) jeśli $\mathscr{K}_{\mathrm{I}}^k = \emptyset$ oraz $\mathscr{K}_{\mathrm{III}}^k \neq \emptyset$, to r odpowiada składowej w_r^k takiej, że $|w_r^k| = \max\limits_{i \in \mathscr{K}_{\mathrm{III}}^k} |w_i^k|$;

(c) jeśli $\mathscr{K}_{\mathrm{I}}^k = \emptyset$, $\mathscr{K}_{\mathrm{III}}^k = \emptyset$ oraz $\mathscr{K}_{\mathrm{II}}^k \neq \emptyset$, to r odpowiada składowej w_r^k takiej, że $|w_r^k| = \max\limits_{i \in \mathscr{K}_{\mathrm{II}}^k} |w_i^k|$.

C (a) Tak samo jak A (a);

(b) jeśli $\mathscr{K}_{\mathrm{I}}^k = \emptyset$, to r odpowiada składowej w_r^k takiej, że $|w_r^k| = \max\limits_{i \in \mathscr{K}_{\mathrm{II}}^k \cup \mathscr{K}_{\mathrm{III}}^k} |w_i^k|$.

D Indeks r odpowiada składowej w_r^k takiej, że $|w_r^k| = \max_{i \in \mathscr{K}^k} |w_i^k|$.

Strategie A, B, C można interpretować w następujący sposób.

Z bazy P^k usuwamy w pierwszej kolejności kolumnę wolną p_r^k, której odpowiada kierunek π_r^k najszybszego spadku funkcji celu taki, że $r \in \mathscr{K}_{\mathrm{I}}^k$. Dopiero jeśli takie kolumny wolne nie istnieją, tzn. $\mathscr{K}_{\mathrm{I}}^k = \emptyset$, to usuwamy z bazy kolumnę ograniczoną p_r^k, której odpowiada kierunek π_r^k najszybszego spadku funkcji celu taki, że $r \in \mathscr{K}_{\mathrm{II}}^k \cup \cup \mathscr{K}_{\mathrm{III}}^k$. W zależności od strategii będzie to najpierw kolumna ograniczona dolna w punkcie x^k (strategia A) lub górna w punkcie x^k (strategia B) lub po prostu ograniczona w punkcie x^k (strategia C).

Strategia D polega na wyborze kolumny p_r^k, której odpowiada kierunek π_r^k najszybszego spadku funkcji celu taki, że $r \in \mathscr{K}^k$. Aby jednak i w tym przypadku metoda była zbieżna, musimy nałożyć dodatkowy warunek:

$$(2.6)\quad \begin{cases} \text{Gdy wybrane } r \text{ nie należy do zbioru } \mathscr{K}_{\mathrm{I}}^k,\ \theta^k = 0 \text{ oraz jeśli } \mathscr{K}_{\mathrm{I}}^k \neq \emptyset, \\ \text{wówczas } r \text{ należy zastąpić } \tilde{r} \text{ takim, że} \\ \qquad |w_{\tilde{r}}^k| = \max_{i \in \mathscr{K}_{\mathrm{I}}^k} |w_i^k|. \end{cases}$$

Celowość wprowadzenia powyższego warunku uzasadniają rozważania przeprowadzone w punkcie 4 (dowód lematu 2).

U w a g a. Wybór indeksu r może być dokonany również według strategii A1, B1, C1, D1 różniących się od A, B, C, D tylko tym, że zamiast wybierać kolumnę p_r^k taką, że odpowiada jej kierunek π_r^k najszybszego spadku funkcji celu, wybiera się p_r^k taką, że odpowiada jej kierunek π_r^k, wzdłuż którego wartość funkcji celu maleje najwięcej.

TWIERDZENIE 1. *Jeśli w punkcie dopuszczalnym x^k nie zachodzi żaden z przypadków* I, II, III, *to x^k jest punktem optymalnym dla zadania* (1.1).

D o w ó d. Fakt, że nie zachodzi żaden z przypadków I, II, III oznacza, że $w_i^k = 0$ dla $i \notin R^k$, $w_i^k \geqslant 0$ dla $i \in R_\beta^k$ oraz $w_i^k \leqslant 0$ dla $i \in R_\alpha^k$.

Pokażemy, że dla wektora współczynników Lagrange'a λ określonego następująco:

$$(2.7)\qquad \lambda_{\beta j} = \begin{cases} w_i^k, & \text{jeśli } (\exists i \in R_\beta^k) p_i^k = e_j, \\ 0 & \text{w przeciwnym przypadku;} \end{cases}$$

$$(2.8)\qquad \lambda_{\alpha j} = \begin{cases} w_i^k, & \text{jeśli } (\exists i \in R_\alpha^k) p_i^k = e_j, \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

spełnione są warunki optymalności Kuhna–Tuckera dla zadania (1.1). Rzeczywiście, z definicji wektora λ oraz z dopuszczalności x^k wynika, że $\lambda_\beta \geqslant 0$, $\lambda_\alpha \leqslant 0$ oraz $\beta \leqslant x^k \leqslant \alpha$. Z zależności

$$z^k = P^k w^k = \sum_{i=1}^{n} w_i^k p_i^k = \sum_{i \in R_\beta^k} w_i^k e_{j_i} + \sum_{i \in R_\alpha^k} w_i^k e_{j_i} = \lambda_\beta + \lambda_\alpha$$

otrzymujemy $z^k - \lambda_\beta - \lambda_\alpha = 0$. Ponadto $\lambda_\beta^T(x^k - \beta) = 0$ oraz $\lambda_\alpha^T(x^k - \alpha) = 0$ z definicji zbiorów R_β^k, R_α^k i z definicji wektorów λ_β, λ_α.

Wobec powyższych rozważań, x^k jest rozwiązaniem optymalnym zadania (1.1). ■

3. Algorytm SUBQ. Obliczenia rozpoczynamy z punktu $x^0 = \beta$, z bazą startową $P^0 = (P^0)^{-1} = I$, $z^0 = d + D\beta$, $w^0 = z^0$, $R^0 = R_\beta^0 = \{1, \ldots, n\}$.

k-ta iteracja algorytmu

K r o k 1. Obliczyć $w^k = (P^k)^{-1} z^k$.

K r o k 2. Jeśli $\mathscr{K}^k = \emptyset$, to x^k jest punktem optymalnym. Optymalną wartością funkcji celu jest $f(x^k) = \frac{1}{2} x^{kT}(d + z^k)$. W przeciwnym razie przejść do kroku 3.

K r o k 3. Wyznaczyć indeks r spośród $i \in \mathscr{K}^k$ według jednej ze strategii zaprezentowanych w punkcie 2. Indeksowi r odpowiada kolumna p_r^k, którą z bazy usuwamy. Obliczyć $q^k = D\pi_r^k$.

K r o k 4. Obliczyć wielkości:

$$\theta_0^k = \begin{cases} \operatorname{sign} w_r^k, & \text{gdy } q^k = 0, \\ w_r^k / (q^{kT} \pi_r^k), & \text{gdy } q^k \neq 0; \end{cases} \tag{3.1}$$

$$\theta_0^k \pi_r^k \text{ oraz } t_0^k \text{ (według wzoru (2.3))}.$$

Jeśli $q^k \neq 0$ oraz $t_0^k > 1$, to $\theta^k = \theta_0^k$. Przejść do kroku 5. Jeśli $q^k \neq 0$ oraz $0 \leqslant t_0^k \leqslant 1$ lub $q^k = 0$, to $\theta^k = t_0^k \theta_0^k$. Przejść do kroku 6.

K r o k 5. Dokonać zmiany bazy wprowadzając do niej zamiast kolumny p_r^k kolumnę wolną $p_r^{k+1} = q^k$. Przejść do kroku 7.

K r o k 6. Dokonać zmiany bazy wprowadzając do niej zamiast kolumny p_r^k kolumnę ograniczoną $p_r^{k+1} = e_l$ taką, że $e_l^T x^{k+1} = \beta_l$ (gdy $t_0^k = t_0^{k\beta}$) lub $e_l^T x^{k+1} = \alpha_l$ (gdy $t_0^k = t_0^{k\alpha}$).

K r o k 7. Przeprowadzić aktualizację x^k, z^k, $(P^k)^{-1}$ oraz zbiorów R^k, R_β^k, R_α^k zgodnie z poniższymi wzorami:

$$x^{k+1} = x^k - \theta^k \pi_r^k; \quad z^{k+1} = z^k - \theta^k q^k; \tag{3.2}$$

$$\pi_r^{k+1} = \frac{1}{(p_r^{k+1})^T \pi_r^k} \pi_r^k; \quad \pi_i^{k+1} = \pi_i^k - \frac{(p_r^{k+1})^T \pi_i^k}{(p_r^{k+1})^T \pi_r^k} \pi_r^k \text{ dla } i \neq r; \tag{3.3}$$

$$R^{k+1} = R(x^{k+1}) = \begin{cases} R^k, & \text{gdy } (p_r^k, p_r^{k+1} \text{ — obie ograniczone) lub} \\ & (p_r^k, p_r^{k+1} \text{ — obie wolne}), \\ R^k - \{r\}, & \text{gdy } p_r^k \text{ — ograniczona}, p_r^{k+1} \text{ — wolna}, \\ R^k \cup \{r\}, & \text{gdy } p_r^k \text{ — wolna}, p_r^{k+1} \text{ — ograniczona} \end{cases} \tag{3.4}$$

oraz

$$R_\beta^{k+1} = R_\beta(x^{k+1}) = \begin{cases} R_\beta^k, & \text{gdy } (p_r^k, p_r^{k+1} \text{ — obie ograniczone dolne)} \\ & \text{lub } (p_r^k \text{ — ograniczona górna lub wolna,} \\ & p_r^{k+1} \text{ — ograniczona górna lub wolna}), \\ R_\beta^k - \{r\}, & \text{gdy } p_r^k \text{ — ograniczona dolna}, p_r^{k+1} \text{ — ogra-} \\ & \text{niczona górna lub wolna}, \\ R_\beta^k \cup \{r\}, & \text{gdy } p_r^k \text{ — ograniczona górna lub wolna,} \\ & p_r^{k+1} \text{ — ograniczona dolna}, \end{cases} \tag{3.5}$$

$$(3.6)\qquad R_\alpha^{k+1} = R_\alpha(x^{k+1}) = \begin{cases} R_\alpha^k, & \text{gdy } (p_r^k, p_r^{k+1} \text{ — obie ograniczone górne) lub } (p_r^k \text{ — ograniczona dolna lub wolna, } p_r^{k+1} \text{ — ograniczona dolna lub wolna),} \\ R_\alpha^k - \{r\}, & \text{gdy } p_r^k \text{ — ograniczona górna, } p_r^{k+1} \text{ — ograniczona dolna lub wolna,} \\ R_\alpha^k \cup \{r\}, & \text{gdy } p_r^k \text{ — ograniczona dolna lub wolna, } p_r^{k+1} \text{ — ograniczona górna.} \end{cases}$$

Przejść do kroku 1.

4. Dowód zbieżności. W niniejszym punkcie przedstawiony jest dowód zbieżności dla strategii A, A1, B, B1, C, C1 i D, D1 przy uwzględnieniu warunku (2.6).

Dla każdego punktu dopuszczalnego x^k zbiór indeksów $R^k = R(x^k)$ wyznacza pewien podzbiór $\mathscr{X}_{R^k}$ zbioru dopuszczalnego $\mathscr{X}$ taki, że

$$\mathscr{X}_{R^k} = \mathscr{X}_{R(x^k)} \overset{\text{df}}{=} \{x \in \mathscr{X}\colon \forall_{i \in R_\beta^k} e_{j_i}^T x = \beta_{j_i} \wedge \forall_{i \in R_\alpha^k} e_{j_i}^T x = \alpha_{j_i}\}.$$

Jeśli $R^k = \emptyset$, to $\mathscr{X}_{R^k} = \mathscr{X}$. Oczywiście, $x^k \in \mathscr{X}_{R(x^k)}$.

TWIERDZENIE 2. *Jeśli w punkcie x^k nie zachodzi przypadek* I, *zachodzi natomiast przypadek* II *lub* III, *to x^k rozwiązuje zadanie*:

$$(4.1)\qquad \min\{f(x) = d^T x + \tfrac{1}{2} x^T D x\colon x \in \mathscr{X}_{R(x^k)}\}, \quad \textit{gdzie } x \in E^n.$$

D o w ó d. Punkt x jest optymalny dla zadania (4.1) wtedy i tylko wtedy gdy istnieje wektor współczynników Lagrange'a $\lambda = \begin{bmatrix} \lambda_\beta \\ \lambda_\alpha \end{bmatrix}$, $\lambda_\beta, \lambda_\alpha \in E^n$ spełniający poniższe warunki Kuhna–Tuckera:

$$\begin{cases} z - \lambda_\beta - \lambda_\alpha = 0; & \beta \leqslant x \leqslant \alpha; \\ \forall_{i \in R_\beta^k} x_{j_i} = \beta_{j_i}, & \forall_{i \in R_\alpha^k} x_{j_i} = \alpha_{j_i}; \\ \lambda_\beta^T(x - \beta) = 0, & \lambda_\alpha^T(x - \alpha) = 0 \end{cases}$$

i taki, że $\lambda_{\beta j_i} \geqslant 0$ dla $i \notin R_\beta^k$ oraz $\lambda_{\alpha j_i} \leqslant 0$ dla $i \notin R_\alpha^k$.

Pokażemy, że warunki te są spełnione w punkcie x^k, jeżeli $\lambda_{\beta j}$ określimy zgodnie z (2.7), $\lambda_{\alpha j}$ zaś zgodnie z (2.8).

Z założenia: $w_i^k = 0$ dla $i \notin R^k = R(x^k)$ wynika, że

$$z^k = P^k w^k = \sum_{i=1}^n w_i^k p_i^k = \sum_{i \in R^k} w_i^k p_i^k = \sum_{i \in R_\beta^k} w_i^k e_{j_i} + \sum_{i \in R_\alpha^k} w_i^k e_{j_i} =$$
$$= \sum_{i \in R_\beta^k} \lambda_{\beta j_i} e_{j_i} + \sum_{i \in R_\alpha^k} \lambda_{\alpha j_i} e_{j_i},$$

czyli $z^k - \lambda_\beta - \lambda_\alpha = 0$. Ponadto, $\beta \leqslant x^k \leqslant \alpha$ oraz $e_{j_i}^T x^k = \beta_{j_i}$ dla $i \in R_\beta^k$, $e_{j_i}^T x^k = \alpha_{j_i}$ dla $i \in R_\alpha^k$, ponieważ $x^k \in \mathscr{X}_{R(x_k)}$.

Z definicji wektorów λ_β, λ_α oraz z definicji zbiorów R_β^k, R_α^k wynika, że $\lambda_{\beta j_i} = 0$

dla $i \notin R_\beta^k$, $\lambda_{\alpha j_i} = 0$ dla $i \notin R_\alpha^k$ oraz $\lambda_\beta^T(x^k - \beta) = 0$, $\lambda_\alpha^T(x^k - \alpha) = 0$. Z tego wynika, że punkt x^k jest rozwiązaniem optymalnym zadania (4.1). ■

Zauważmy, że jeśli w punkcie x^k mamy $R^k = \emptyset$ oraz nie zachodzi przypadek I, to x^k (na mocy twierdzenia 2) minimalizuje funkcję celu na zbiorze $\mathscr{X}_{R^k} = \mathscr{X}$, czyli jest punktem optymalnym dla zadania (1.1). Algorytm SUBQ ma taką własność, że wartość funkcji celu nie rośnie, gdy przechodzi się do kolejnej iteracji. Wobec tego, że każdej iteracji odpowiada pewien zbiór R^k (a więc $\mathscr{X}_{R(x^k)}$), można przejść tylko do „niegorszego" $\mathscr{X}_{R^k}$ (w sensie malenia kryterium).

Liczba podzbiorów $\mathscr{X}_{R^k}$ jest skończona.

Gdybyśmy potrafili udowodnić, że ma miejsce następujący ciąg zdarzeń:

{przypadek II lub III — malenie wartości funkcji celu — przypadek II lub III},

to ciąg ten byłby skończony. Każdemu kolejnemu przypadkowi II lub III odpowiadałby bowiem wówczas „lepszy" zbiór $\mathscr{X}_{R^k}$.

Udowodnijmy następujące dwa twierdzenia:

TWIERDZENIE 3. *Gdy zachodzi przypadek* I, *wówczas przypadek* II *lub* III *zajdzie po skończonej liczbie iteracji lub zostanie osiągnięte optimum.*

WNIOSEK. *Jeśli* $R^k = \emptyset$ *i zachodzi przypadek* I, *to po skończonej liczbie* η *iteracji otrzymamy zbiór* $R^{k+\eta} \neq \emptyset$ *lub optimum.*

TWIERDZENIE 4. *Gdy zachodzi przypadek* II *lub* III, *wówczas wartość funkcji celu maleje lub w następnej iteracji zajdzie przypadek* I *lub osiągane jest optimum.*

Dowód twierdzenia 3. Jeśli zachodzi przypadek I, czyli $w_i^k \neq 0$ dla pewnego $i \notin R^k$, to istnieją dwie możliwości:

Ia) $\theta^k = \theta_0^k$,

Ib) $\theta^k = t_0^k \theta_0^k$,

gdzie t_0^k wyznaczone jest zgodnie ze wzorem (2.3).

Przypadek Ib obejmuje sytuację, w której z bazy usuwana jest kolumna wolna p_r^k a na jej miejsce wprowadzana jest kolumna ograniczona $p_r^{k+1} = e_l$ taka, że $e_l^T x^{k+1} = \beta_l$ (czyli dolna) lub $e_l^T x^{k+1} = \alpha_l$ (czyli górna).

Niech s oznacza liczbę wolnych kolumn w bazie. Gdy zachodzi przypadek Ib, wówczas s maleje o 1. Przypadek Ib może więc zajść kolejno co najwyżej s razy. Zauważmy, że gdy $s = 0$, tzn. wszystkie kolumny bazy są ograniczone, wówczas w następnej iteracji musi zajść przypadek II lub III bądź optimum zostało osiągnięte.

Przypadek Ia obejmuje sytuację, w której z bazy usuwana jest kolumna wolna p_r^k a na jej miejsce wprowadzana jest kolumna wolna $p_r^{k+1} = q^k$. W przypadku tym s nie zmienia się. Okazuje się jednak, że przypadek Ia może zajść kolejno również co najwyżej s razy. Wynika to z poniższego lematu.

LEMAT 1. *Jeżeli zachodzi przypadek* Ia *i w bazie wymieniamy r-tą kolumnę* p_r^k, *to r-ta składowa* w_r^{k+1} *wektora* w^{k+1} *ma wartość zero. Ponadto, jeśli w następnych iteracjach zachodzą po sobie kolejno przypadki* Ia, *to r-ta składowa aktualnego wektora w pozostaje zerem.*

D o w ó d. Ze wzorów (3.2) i (3.3) można wyprowadzić następujące zależności:

$$(4.2)\qquad w_i^{k+1} = (\pi_i^{k+1})^T z^{k+1} =$$

$$= \begin{cases} w_i^k(\theta^k) & \text{dla} \quad i \neq r,\ \theta^k = \theta_0^k, \\ w_i^k(\theta^k) - \dfrac{(p_r^{k+1})^T \pi_i^k}{(p_r^{k+1})^T \pi_r^k} w_r^k(\theta^k) & \text{dla} \quad i \neq r,\ \theta^k = t_0^k \theta_0^k, \\ 0 & \text{dla} \quad i = r,\ \theta^k = \theta_0^k, \\ \dfrac{1}{(p_r^{k+1})^T \pi_r^k} w_r^k(\theta^k) & \text{dla} \quad i = r,\ \theta^k = t_0^k \theta_0^k, \end{cases}$$

gdzie $w_i^k(\theta) = w_i^k - \theta q^{kT} \pi_i^k$.

Gdy zachodzi przypadek Ia, czyli $\theta^k = \theta_0^k$, wówczas z powyższych wzorów wynika, że $w_r^{k+1} = 0$.

Jeśli w bazie P^k wymieniamy r-tą kolumnę p_r^k na $p_r^{k+1} = q^k \overset{\mathrm{df}}{=} D\pi_r^k$ (jak ma to miejsce w przypadku Ia), to korzystając ze wzorów (3.3) otrzymujemy, że dla $i \neq r$:

$$(\pi_i^{k+1})^T D\pi_r^{k+1} = (\pi_i^{k+1})^T \frac{D\pi_r^k}{(p_r^{k+1})^T \pi_r^k} = \frac{(\pi_i^{k+1})^T p_r^{k+1}}{(p_r^{k+1})^T \pi_r^k} = 0.$$

Pokażemy ponadto, że jeśli przez kolejnych η iteracji nadal wprowadzamy do bazy tylko kolumny wolne, to zawsze zachodzi $(\pi_i^{k+j})^T D\pi_r^{k+j} = 0$ dla $i \neq r$ oraz $w_r^{k+j} = 0$ dla $j = 1, \ldots, \eta$. Rzeczywiście, pokazaliśmy, że dla $j = 1$ twierdzenie jest prawdziwe. Załóżmy, że dla $j-1$ $(j \leqslant \eta)$ zachodzi $(\pi_i^{k+j-1})^T D\pi_r^{k+j-1} = 0$ dla $i \neq r$ oraz $w_r^{k+j-1} = 0$.

Jeśli w punkcie x^{k+j} usuniemy z bazy P^{k+j-1} kolumnę p_l^{k+j-1} a w zamian wprowadzamy kolumny $p_l^{k+j} = q^{k+j-1} \overset{\mathrm{df}}{=} D\pi_l^{k+j-1}$, to $l \neq r$ (bo $w_r^{k+j-1} = 0$) oraz $(q^{k+j-1})^T \pi_r^{k+j-1} = (\pi_l^{k+j-1})^T D\pi_r^{k+j-1} = 0$.

Z tego wynika

$$w_r^{k+j} = w_r^{k+j-1} - \theta^{k+j-1} (q^{k+j-1})^T \pi_r^{k+j-1} = 0,$$

$$\pi_r^{k+j} = \pi_r^{k+j-1} - \frac{(p_l^{k+j})^T \pi_r^{k+j-1}}{(p_l^{k+j})^T \pi_l^{k+j-1}} \pi_l^{k+j-1} = \pi_r^{k+j-1}$$

oraz

$$\underset{i \neq r}{\forall} (\pi_i^{k+j})^T D\pi_r^{k+j} = \left((\pi_i^{k+j-1})^T - \frac{(p_l^{k+j})^T \pi_i^{k+j-1}}{(p_l^{k+j})^T \pi_l^{k+j-1}} (\pi_l^{k+j-1})^T \right) D\pi_r^{k+j-1} =$$

$$= (\pi_i^{k+j-1})^T D\pi_r^{k+j-1} - \frac{(p_l^{k+j})^T \pi_i^{k+j-1}}{(p_l^{k+j})^T \pi_l^{k+j-1}} (\pi_l^{k+j-1})^T D\pi_r^{k+j-1} =$$

$$= 0. \ \blacksquare$$

Z lematu 1 wynika, że przypadek Ia może zajść kolejno co najwyżej s razy. Jeśli następnie zachodzi przypadek Ib, to s maleje o 1. Dlatego przypadek I może zajść kolejno skończoną liczbę razy. Twierdzenie 3 zostało tym samym udowodnione.

D o w ó d t w i e r d z e n i a 4. Jeśli zachodzi przypadek II lub III, czyli

$$\left(\underset{i \notin R^k}{\forall} w_i^k = 0 \wedge \underset{i \in R_\beta^k}{\exists} w_i^k < 0 \right) \quad \text{lub} \quad \left(\underset{i \notin R^k}{\forall} w_i^k = 0 \wedge \underset{i \in R_\alpha^k}{\exists} w_i^k > 0 \right),$$

to istnieją cztery możliwości:

IIa) lub IIIa): $\theta^k = \theta_0^k$,

IIb) lub IIIb): $\theta^k = t_0^k \theta_0^k$,

gdzie t_0^k jest wyznaczone zgodnie ze wzorem (2.3).

Przypadek IIa (IIIa) obejmuje sytuację, w której z bazy usuwana jest kolumna ograniczona dolna (górna) p_r^k a na jej miejsce wprowadzana jest kolumna wolna $p_r^{k+1} = q^k \neq 0$. Wartość funkcji celu w nowym punkcie x^{k+1} można przedstawić jako:

$$f(x^{k+1}) = f(x^k) - \tfrac{1}{2}(\pi_r^{kT} D \pi_r^k)(\theta_0^k)^2,$$

przy czym $\theta_0^k = \dfrac{w_r^k}{q^{kT}\pi_r^k} \neq 0$; $\pi_r^{kT} D \pi_r^k > 0$ (na mocy własności (W3), § 5, rozdz. 1, [1], bo $q^k \neq 0$). Zatem w tym przypadku wartość funkcji celu zawsze maleje, czyli $f(x^{k+1}) < f(x^k)$.

Przypadek IIb (IIIb) obejmuje sytuację, w której z bazy usuwana jest kolumna ograniczona dolna (górna) p_r^k, na jej miejsce zaś wprowadzana jest kolumna ograniczona dolna lub górna $p_r^{k+1} = e_l$.

Wartość funkcji celu w nowym punkcie x^{k+1} można przedstawić jako

$$f(x^{k+1}) = \begin{cases} f(x^k) - \tfrac{1}{2}(\pi_r^{kT} D \pi_r^k) t_0^k (\theta_0^k)^2 (2 - t_0^k), & \text{gdy} \quad q^k \neq 0, \\ f(x^k) - t_0^k \theta_0^k w_r^k, & \text{gdy} \quad q^k = 0, \end{cases}$$

przy czym tak samo jak w przypadku IIa (IIIa), jeśli $q^k \neq 0$, to

$$\theta_0^k = \frac{w_r^k}{q^{kT}\pi_r^k} \neq 0, \qquad \pi_r^{kT} D \pi_r^k > 0,$$

oraz jeśli $q^k = 0$, to

$$\theta_0^k = \operatorname{sign} w_r^k \neq 0.$$

Przypadek IIb (IIIb) zachodzi, gdy $q^k \neq 0$ oraz $0 \leqslant t_0^k \leqslant 1$ lub gdy $q^k = 0$. Dlatego wartość funkcji celu maleje wtedy i tylko wtedy, gdy $t_0^k > 0$.

Aby zakończyć dowód twierdzenia 4, wystarczy udowodnić poniższy lemat.

LEMAT 2. *Jeśli zachodzi przypadek* IIb *lub* IIIb *i wartość funkcji celu nie maleje, czyli* $\theta^k = t_0^k \theta_0^k = 0$, *to w następnej iteracji musi zajść przypadek* I, *jeżeli nie zostało osiągnięte optimum.*

D o w ó d. Wykorzystajmy zależności (4.2) uwzględniając fakt, że $\theta^k = t_0^k \theta_0^k = 0$ oraz $p_r^{k+1} = e_l$:

$$w_i^{k+1} = \begin{cases} w_i^k - \dfrac{e_l^T \pi_i^k}{e_l^T \pi_r^k} w_r^k & \text{dla} \quad i \neq r, \\ \dfrac{1}{e_l^T \pi_r^k} w_r^k & \text{dla} \quad i = r. \end{cases}$$

Zachodzi $R^{k+1} = R^k$.

Pokażemy, że istnieje $i \notin R^{k+1} = R^k$ takie, że $w_i^{k+1} \neq 0$. W tym celu przedstawmy kolumnę $p_r^{k+1} = e_l$ jako kombinację liniową kolumn bazy P^k. Wówczas $e_l \notin P^k$ oraz

$$e_l = \sum_{i=1}^{n} \xi_i p_i^k. \tag{4.3}$$

Kolumna e_l nie może być przedstawiona jako kombinacja liniowa tylko kolumn jednostkowych. Musi zatem istnieć $i \notin R^k$ takie, że $\xi_i \neq 0$. Mnożąc stronami równość (4.3) przez π_j^{kT} dla $j = 1, \ldots, n$ otrzymujemy, że $\xi_j = e_l^T \pi_j^k$ dla $j = 1, \ldots, n$. Z tego oraz ponieważ dla strategii A, A1, B, B1, C, C1 przypadek II (III) może zajść tylko wtedy, kiedy $w_i^k = 0$ dla $i \notin R^k$, mamy

$$\forall_{i \notin R^k} \left(i \neq r \wedge w_i^{k+1} = -\frac{e_l^T \pi_i^k}{e_l^T \pi_r^k} w_r^k \right) \quad \text{oraz} \quad \exists_{i \notin R^k = R^{k+1}} \; w_i^{k+1} \neq 0.$$

W następnej iteracji musi więc zajść przypadek I. Dla strategii D i D1, dzięki uwzględnieniu dodatkowego warunku (2.6), w następnej iteracji musi również zajść przypadek I. ■

Lemat 2 kończy dowód twierdzenia 4.

Z dowodów twierdzeń 3 i 4 wynika, że jedyną możliwością, aby wartość funkcji celu nie malała, jest ciąg zdarzeń:

$$\{\text{przypadek IIb lub IIIb}, \theta^k = 0 \text{ — przypadek Ib}, \theta^{k+1} = 0 \text{ — } \ldots \text{ — przypadek Ib}, \theta^{k+j} = 0 \text{ — przypadek IIb lub IIIb}, \theta^{k+j+1} = 0 \text{ — } \ldots\}.$$

Taki ciąg jest skończony, ponieważ:

1. Przypadek Ib może zajść kolejno co najwyżej s razy (gdzie s jest liczbą kolumn wolnych w bazie), przy czym za każdym razem s maleje o 1;

2. gdy zachodzi przypadek IIb lub IIIb, wówczas $R^{k+1} = R^k$, czyli s nie zmienia się;

3. jeśli otrzymamy $s = 0$, czyli $R^k = \{1, \ldots, n\}$, to w następnej iteracji przypadek I nie jest możliwy; musi zajść przypadek II lub III i $\theta^k \neq 0$, jeżeli optimum nie zostało osiągnięte.

Powyższe rozważania kończą dowód zbieżności algorytmu SUBQ.

5. Zakończenie. Na zakończenie warto wspomnieć o korzyściach płynących ze stosowania algorytmu SUBQ w celu rozwiązania zadań postaci (1.1).

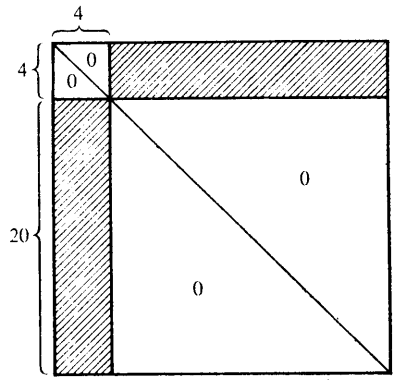
Gdybyśmy na przykład rozwiązywali zadanie (1.1) algorytmem Lemkego [3], musielibyśmy sprowadzić warunki $\beta \leqslant x \leqslant \alpha$ do postaci $A^T x \leqslant d_0$, gdzie $A = [I, -I]$, $d_0 \begin{bmatrix} \alpha \\ -\beta \end{bmatrix}$. Aby można było zadanie (1.1) rozwiązać tym algorytmem, macierz D musi być *dodatnio* określona. Algorytm SUBQ wymaga jedynie, aby macierz D była *nieujemnie określona.*

Testy programu opartego na algorytmie SUBQ, przeprowadzone na emc IBM 360/50, potwierdzają celowość stosowania tej metody w przypadku zadania postaci (1.1).

W celu zilustrowania powyższego stwierdzenia przedstawmy wyniki otrzymane

dla zadania praktycznego rozwiązanego dwiema metodami, tj. programami opartymi na algorytmie Lemkego oraz na SUBQ.

Wymiar zadania wynosi 24; dodatnio określona macierz $D = D^T$ jest postaci:



W przypadku stosowania metody Lemkego wynik został otrzymany po czasie operacyjnym wynoszącym 1 min 18,62 s. Stosując algorytm SUBQ wynik otrzymaliśmy w czasie operacyjnym o prawie 30 s krótszym (dla rozmaitych strategii wyboru indeksu r czasy były różne, odchylenia jednak niewielkie).

Z porównania czasów oraz liczby iteracji dla przeprowadzonych testów wynika, że na przebieg algorytmu ma wpływ nie tylko stosowanie odpowiedniej strategii A, B, C lub D, lecz także właściwy wybór punktu startowego.

Najlepiej widoczne jest to na trywialnym przykładzie liniowej funkcji celu, tzn. gdy $D = 0$. Rozwiązaniem optymalnym jest wówczas któryś z wierzchołków zbioru dopuszczalnego $\mathscr{X} = \{x\colon \beta \leqslant x \leqslant \alpha\}$. Startując z tego wierzchołka, tj. z punktu $x^0 = \gamma$ takiego, że

$$\gamma_i = \begin{cases} \beta_i, & \text{gdy} \quad d_i > 0, \\ \alpha_i, & \text{gdy} \quad d_i < 0, \end{cases}$$

wykonuje się zawsze tylko jedną iterację. Startując z punktu $x^0 = \beta$ w najgorszym razie będzie ich n.

W przypadku kwadratowej funkcji celu wybór punktu startowego nie jest tak prosty. Najkorzystniej jest, oczywiście, za punkt startowy przyjąć taki punkt, który rokuje jak najszybsze uzyskanie wyników. W tym celu należałoby zbadać przebieg funkcji celu lub przynajmniej obliczyć $z_i(\beta)$, $z_i(\alpha)$ dla $i = 1, \ldots, n$, gdzie $z_i(x) \overset{\mathrm{df}}{=} \left.\dfrac{\partial f}{\partial x_i}\right|_x$.

Przyjmijmy za punkt startowy punkt $x^0 = \gamma$ taki, że

$$\gamma_i = \begin{cases} \alpha_i, & \text{gdy } i \in \mathfrak{N}(\beta) \cap \mathfrak{N}(\alpha), \\ \beta_i, & \text{gdy } i \notin \mathfrak{N}(\beta) \cup \mathfrak{N}(\alpha), \\ \beta_i, & \text{gdy } i \in \mathfrak{N}(\beta) \setminus \mathfrak{N}(\alpha) \text{ oraz } |z_i(\beta)| \leqslant |z_i(\alpha)|, \\ \alpha_i, & \text{gdy } i \in \mathfrak{N}(\beta) \setminus \mathfrak{N}(\alpha) \text{ oraz } |z_i(\beta)| > |z_i(\alpha)|, \end{cases}$$

gdzie

$$\mathfrak{N}(\beta) \stackrel{\mathrm{df}}{=} \{i\colon z_i(\beta) \leqslant 0\}, \qquad \mathfrak{N}(\alpha) \stackrel{\mathrm{df}}{=} \{i\colon z_i(\alpha) \leqslant 0\}.$$

(Takie rozszerzenie na przypadek kwadratowej funkcji celu jest poprawne, bo gdy $D = 0$, wówczas otrzymujemy opisany powyżej sposób wyboru punktu startowego dla przypadku liniowej funkcji celu.)

Przeprowadzone dla opisanego wcześniej zadania testy uwzględniające strategie A, B, C, D oraz różne sposoby wyboru punktu startowego potwierdzają słuszność wyboru punktu γ za punkt startowy. W przypadku, gdy za punkt startowy przyjmowaliśmy $x^0 = \beta$, wynik był otrzymywany po 25 iteracjach. Gdy punktem startowym był punkt $x^0 = \gamma$, wynik był otrzymywany po 20 iteracjach. Również wyniki otrzymane dla innego zadania praktycznego (o 9 zmiennych i dodatnio określonej macierzy D) potwierdzają słuszność wyboru punktu γ za punkt startowy. Rozwiązanie optymalne zostało wówczas otrzymane po 6 iteracjach bez względu na zastosowaną strategię A, B, C lub D. W przypadku, gdy za punkt startowy przyjmowaliśmy β, rozwiązanie optymalne było otrzymane po 11 iteracjach bez względu na zastosowaną strategię A, B, C, D. Gdy punktem startowym był punkt $x^0 = \alpha$, wówczas wprawdzie dla strategii A, C, D otrzymaliśmy wynik po 8 iteracjach, ale strategia B dawała wynik po 16 iteracjach.

Uwaga. Metodę SUBQ można łatwo rozszerzyć na przypadek, gdy tylko pewne ograniczenia są postaci $\beta_i \leqslant x_i \leqslant \alpha_i$, pozostałe zaś są typu $x_i \geqslant \beta_i$ lub $x_i \leqslant \alpha_i$. Wystarczy wówczas założyć, że funkcja celu przyjmuje w zbiorze dopuszczalnym minimum skończone. Punktem startowym jest punkt x^0 taki, że

$$x_i^0 = \begin{cases} \alpha_i & \text{dla ograniczeń typu } x_i \leqslant \alpha_i, \\ \beta_i & \text{dla ograniczeń typu } x_i \geqslant \beta_i, \\ \gamma_i & \text{dla ograniczeń typu } \beta_i \leqslant x_i \leqslant \alpha_i. \end{cases}$$

Literatura

[1] W. Dubnicki i K. Zorychta, *Metody programowania wypukłego*, Wydawnictwa Uniwersytetu Warszawskiego, 1972.

[2] G. Hille, *Analityczna i geometryczna interpretacja metody Lemkego dla zadań programowania kwadratowego*, Pr. IBS PAN nr 4.

[3] C. E. Lemke, *A method of solution for quadratic programming*, Manag. Sci. 8 (1962B), str. 442–453.